SOURCE AND RELIABILITY STATEMENT FOR SIPP
WAVE 7

SOURCE OF DATA

The data were collected in the seventh interview wave of the 1984 panel of the Survey of Income and Program Participation (SIPP). The SIPP universe is the noninstitutionalized resident population living in the United States. This population includes persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents, were not eligible to be in the survey.

Similarly, United States citizens residing abroad were not eligible to be in the survey. Foreign visitors who work or attend school in this country and their families were eligible; all others were not eligible to be in the survey. With the exceptions noted above, persons who were at least 15 years of age at the time of the interview were eligible to be in the survey.

The 1984 panel SIPP sample is located in 174 areas comprising 450 counties (including one partial county) and independent cities. Within these areas, clusters of 2 to 4 living quarters (LQs) were systematically selected from lists of addresses prepared for the 1970 decennial census to form the bulk of the sample. To account for LQs built within each of the sample areas after the 1970 census, a sample was drawn of permits issued for construction of residential LQs through March 1983. In jurisdictions that do not issue building permits, small land areas were sampled and the LQs within were listed by field personnel and then subsampled. In addition, sample LQs were selected from supplemental frames that included mobile home parks and new construction for which permits were issued prior to January 1, 1970, but for which construction was not completed until after April 1, 1970.

Approximately 26,000 living quarters were originally designated for the sample. For Wave 1, interviews were obtained from the occupants of about 19,900 of the 26,000 designated living quarters. Most of the remaining 6,100 living quarters were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, approximately 1,000 of the 6,100 living quarters were not interviewed because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 95 percent of all eligible living quarters participated in wave 1 of the survey.

For the subsequent waves, only original sample persons (those interviewed in the first wave) and persons living with them were eligible to be interviewed. With certain restrictions, original sample persons were to be followed if they moved to a new address. All noninterviewed households from Wave 1 were automatically designated as noninterviews for all subsequent waves. When original sample persons moved without leaving a forwarding address or moved to extremely remote parts of the country, additional noninterviews resulted.

Sample households within a given panel are divided into four subsamples of nearly equal size. These subsamples are called rotation groups, denoted R (R = 1, 2, 3, or 4), and one rotation group is interviewed each month. Each household in the sample was scheduled to be interviewed at 4 month intervals over a period of 2 1/2 years beginning in October 1983. The reference period for the questions is the 4-month period preceding the interview. In general, one cycle of four interviews covering the entire sample, using the same questionnaire, is called a wave.

The Wave 7 public use file includes core data and supplemental (topical module) data. Core questions are repeated at each interview over the life of the panel. Topical modules include questions which are not asked every month. The Wave 7 topical module covers (1) Assets and Liabilities, (2) Pension Plan Coverage and (3) Real Estate Property and Vehicles.

Table 1 indicates the reference months and interview month for the collection of data from each rotation group for Wave 7. For example, rotation group 2 was interviewed in November 1985 and data for the reference months July 1985 through October 1985 were collected.

Table 1. Reference Months for Each Interview Month - Wave 7

Month of Interview	Rotation	Reference Period Second Quarter (1985)			Third Quarter (1985)			Fourth Quarter (1985)	
		Apr	May	Jun	Jul	Aug	Sept	Oct	Nov
September	4		X	X	X	X			
October	1			X	X	X	X		
November	2				X	X	X	X	
December	3					X	X	X	X

The estimation procedure used to derive SIPP person weights involves several stages of weight adjustments. These include determining the base weight, adjusting for movers and noninterviews, adjusting to account for the SIPP sample areas not having the same population distribution as the strata from which they were selected and adjusting persons' weights to bring sample estimates into agreement with independent population estimates.

Each person received a base weight equal to the inverse of his/her probability of selection. The SIPP base weight W indicates that each SIPP sample person represents approximately W persons in the SIPP universe. Due to funding difficulties, a sample cut of 17.8 percent was implemented in March 1985. Each rotation group was reduced by about 850 interviewed housing units.

Both self-representing (SR) PSUs and nonself-representing (NSR) PSUs were subject to the cut. In some instances, the base weight was adjusted to reflect subsampling done in the field. For each subsequent interview, each person received a base weight that accounted for following movers.

A noninterview adjustment factor was applied to the weight of each interviewed person to account for persons in occupied living quarters who were eligible for the sample but were not interviewed. (Individual nonresponse within partially interviewed households was treated with imputation. No special adjustment was made for noninterviews in group quarters.)

A first stage ratio estimate factor was applied to each interviewed person's weight to account for the SIPP NSR sample areas not having the same population distribution as the strata from which they were selected. In particular, the first stage ratio estimate factors ensure proportional representation by race and by metropolitan and nonmetropolitan residence defined as of June 1981.

An additional stage of adjustment to persons' weights was performed to bring the sample estimates into agreement with independent monthly estimates of the civilian (and some military) noninstitutional population of the United States by age, race, and sex. These independent estimates were based on statistics from the 1980 Decennial Census of Population; statistics on births, deaths, immigration, and emigration; and statistics on the strength of the Armed Forces. Weights were further adjusted so that sample estimates would agree with special Current Population Survey (CPS) estimates of the prevalence of different types of householders (married, single with relatives or single without relatives by sex and race) and different relationships to householders (spouse or other). Also, husbands and wives were assigned equal weights. As a result of these adjustments, the following types of consistency are attained by race and sex on a monthly basis:

1. The sum of weights of civilian (and some military) noninstitutionalized persons agrees with independent estimates by age groups.

2. The sum of weights of civilian (and some military) noninstitutionalized persons is within a close tolerance of special CPS estimates by householder type and relationship to householder. (The special CPS estimates are similar but not identical to the monthly CPS estimates.)

3. Husbands and wives living together have equal weights. Thus, if a characteristic is necessarily shared by a husband and wife (such as size of family), then the sample estimate of the number of husbands with the characteristic will agree with the corresponding estimate for wives.

Use of Weights. Each household and each person within each household on the Wave 7 tape has five weights. Four of these weights are reference month specific and therefore can be used only to form reference month estimates. To form an estimate for a particular month, use the reference month weight for the month, summing over all persons or households with the characteristic of interest whose reference period includes that month. Multiply the sum by a factor to account for the number of rotations contributing data for the month. This factor equals four divided by the number of rotations contributing data for the month. For example, July data are only available from rotations 1, 2, and 4 (see Table 1), so a factor of 4/3 must be applied. August data are available from all four rotations, so a factor of 4/4 = 1 must be applied. Reference month estimates can be averaged to form estimates of monthly averages over some period of time. For example, using the proper weights, one can estimate the monthly average number of households in a specified income range over October and November 1985.

The remaining weight is interview month specific. This weight can be used to form estimates that specifically refer to the interview month (e.g., total persons currently looking for work), as well as estimates referring to the time period including the interview month and all previous months (e.g., total persons who have ever served in the military). There is no weight for characteristics that involve a person's or household's status over two or more months (e.g., number of households with a 50 percent increase in income between October and November 1985).

When estimates for all months except August are constructed from Wave 7 data, factors greater than 1 must be applied. However, when the wave 7 core data are used in conjunction with the wave 6 and wave 8 core data, data from all four rotations will be available for April through December, and the factors will equal 1 for those months.

To estimate monthly averages of a given measure (e.g., total, mean) over a number of consecutive months, sum the monthly estimates and divide by the number of months.

Producing Estimates for Census Regions. The total estimate for a region is the sum of the state estimates in that region. However, one of the groups of states formed for confidentiality reasons crosses regional boundaries. This group consists of South Dakota (Midwest Region), Idaho (West Region), New Mexico (West Region), and Wyoming (West Region). To compute the total estimate for the Midwest Region, a factor of 0.203 should be applied to the above group's total estimate and added to the sum of the other state estimates in the Midwest Region. For the West Region, a factor of 0.797 should be applied to the above group's total estimate and added to the sum of the other states in the West Region.

Estimates from this sample for individual states are subject to very high variance and are not recommended. The state codes on the file are primarily of use for linking respondent characteristics with appropriate contextual variables (e.g., state-specific welfare criteria) and for tabulating data by user-defined groupings of states.

Producing Estimates for the Metropolitan Population. For 15 states in the SIPP sample, metropolitan or nonmetropolitan residence is identified (Variable H*-METRO, characters 94, 382, 670, and 958). In 21 additional states, where the nonmetropolitan population in the sample was small enough to present a disclosure risk, a fraction of the metropolitan sample was recoded so as to be indistinguishable from nonmetropolitan cases (H*-METRO=2). In these states, therefore, the cases coded as metropolitan (H*-METRO=1) represent only a subsample of that population.

In producing state estimates for a metropolitan characteristic, multiply the individual, family, or household weights by the metropolitan inflation factor for that state presented in table 4. (This inflation factor compensates for the subsampling of the metropolitan population and is 1.0 for the states with complete identification of the metropolitan population.)

The same procedure applies when creating estimates for particular identified MSA's or CMSA's--apply the factor appropriate to the state. For multi-state MSA's, use the factor appropriate to each state part. For example, to tabulate data for the Washington, DC-MD-VA MSA, apply the Virginia factor of 1.0778 to weights for residents of the Virginia part of the MSA; Maryland and Washington, DC residents require no modification to the weights (i.e., their factors equal 1.0).

In producing regional or national estimates of the metropolitan population, it is also necessary to compensate for the fact that no metropolitan subsample is identified within two states (Maine and Iowa) and one state-group (Mississippi-West Virginia). There were no metropolitan areas sampled in South Dakota-Idaho-New Mexico-Wyoming. Therefore, a different factor for regional and national estimates is in the right-hand column of table 4. The results of regional and national tabulations of the metropolitan population will be biased slightly. However, less than one-half of one percent of the metropolitan population is not represented.

Producing Estimates for the Nonmetropolitan Population. State, regional, and national estimates of the nonmetropolitan population cannot be computed directly, except for the 15 states where the factor in table 4 is 1.0. In all other states, the cases identified as not in the metropolitan subsample (METRO=2) are a mixture of nonmetropolitan and metropolitan households. Only an indirect method of estimates is available: first compute an estimate for the total population, then subtract the estimate for the metropolitan population. The results of these tabulations will be slightly biased.

RELIABILITY OF THE ESTIMATES

SIPP estimates in this report are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: nonsampling and sampling. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error. Found below are descriptions of sources of SIPP nonsampling error, followed by a discussion of sampling error, its estimation, and its use in data analysis.

Nonsampling Variability. Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness on the part of the respondents to provide correct information, inability to recall information, errors made in collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, biases resulting from the differing recall periods caused by the rotation pattern used, and failure to represent all units within the sample (undercoverage). Quality control and edit procedures were used to reduce errors made by respondents, coders, and interviewers.

Undercoverage in SIPP results from missed living quarters and missed persons within sample households. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for blacks than for nonblacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have different characteristics from those of the interviewed persons in the same age-race-sex group. Further, the independent population controls used have not been adjusted for undercoverage in the decennial census.

The following table summarizes information on household nonresponse for the interview months used to produce this report.

Sample Size, by Month and Interview Status

Month	Household Units Eligible			
	Total	Inter-viewed	Not Inter-viewed	Non-Response Rate
Sept 1985	4,700	3,700	1,000	22
Oct 1985	4,700	3,800	900	20
Nov 1985	4,700	3,700	900	20
Dec 1985	4,800	3,700	1,100	22

Due to rounding of all numbers at 100, there are some inconsistencies. The percentage was calculated using unrounded numbers.

Some respondents do not respond to some of the questions. Therefore, the overall nonresponse rate for some items, such as income and money-related items is higher than the nonresponse rates in the above table. The Bureau has used complex techniques to adjust the weights for nonresponse, but the success of these techniques in avoiding bias is unknown.

Comparability with other statistics. Caution should be exercised when comparing data from this file with data from other SIPP products or with data from other surveys. The comparability problems are caused by the seasonal patterns for many characteristics and by different nonsampling errors.

Sampling variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Hypothesis Testing. Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses tested are 1) the population parameters are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the parameters are different when, in fact, they are identical.

To perform the most common test, let X_A and X_B be sample estimates of two parameters of interest. A subsequent section explains how to derive a standard error on the difference $X_A - X_B$. Let that standard error be s_{DIFF}. Compute the ratio $R = (X_A - X_B)/s_{DIFF}$. If this ratio is between -1.6 and $+1.6$, no conclusion about the parameters is justified at the 10 percent significance level. If, on the other hand, this ratio is smaller than -1.6 or larger than $+1.6$, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the parameters are different. Of course, sometimes this conclusion will be wrong. When the parameters are, in fact, the same, there is a 10 percent chance of concluding that they are different.

Note when using small estimates. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 200,000. Nonsampling error can occasionally occur in one of the small number of cases used in the estimate, causing large relative error in that particular estimate. Also, care must be taken in the interpretation of small differences. Even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard Error Parameters and Tables and Their Use. To derive standard errors that would be applicable to a wide variety of statistics and could be prepared at a moderate cost, a number of approximations were required. All statistics do not have the same variance behavior; statistics with similar variance behavior were grouped together. Most of the SIPP statistics have greater variance than those obtained through a simple random sample because clusters of living quarters are sampled for SIPP. Two parameters (denoted "a" and "b") were developed to quantify these increases in variance. These "a" and "b" parameters are used in estimating standard errors of survey estimates. The "a" and "b" parameters vary by type of estimate and by subgroup to which the estimate applies. Table 3 provides base "a" and "b" parameters for various subgroups and types of estimates. For SIPP wave 7 core and topical module characteristics, factors for each of the single reference months, May 1985 through November 1985, are provided. The factor multiplied by the base parameters for a given subgroup and type of estimate gives the "a" and "b" parameters for that subgroup and estimate type in the chosen time period. For example, the base "a" and "b" parameters for total income of households are -0.0000993 and 8582, respectively. The factor for May 1985 is 4, so that "a" and "b" parameters for total household income in May 1985 are -0.0003972 and 34,328, respectively.

The "a" and "b" parameters may be used directly to calculate the standard error for estimated numbers and percentages. Because the actual variance behavior was not identical for all statistics within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific statistic. Methods for using these parameters for direct computation of standard errors are given in the following sections.

The user can create far more types of estimates than standard errors are provided for here. Procedures for calculating standard errors for the types of estimates most commonly used are described below. Note specifically that these procedures apply only to reference month estimates or averages of reference month estimates. Refer to the section "Use of Weights" for a detailed discussion of construction of estimates.

Stratum codes and half sample codes are included on the tape to enable the user to compute the variances directly from the data by methods such as balanced repeated replications (BRR). William G. Cochran provides a list of references discussing the application of this technique.[1]

Standard errors of estimated numbers. The approximate standard error of an estimated number can be obtained by using formula (1).

$$s_x = \sqrt{ax^2 + bx} \qquad (1)$$

Here x is the size of the estimate and "a" and "b" are the parameters associated with the particular type of characteristic for the appropriate reference period. Note that this method should not be applied to dollar values.

Illustration of the computation of the standard error of an estimated number. Suppose that SIPP estimates for October 1985 show that there were 472,000 HHs outside metropolitan areas with monthly household income above $6,000. Then the appropriate "a" and "b" parameters and factor to use in calculating a standard error for the estimate are obtained from table 3. They are a = −0.0000993, b = 8,582 and a factor of 2 for October.

Using formula (1), the approximate standard error is

$$\sqrt{(-0.001986)(472{,}000)^2 + (17{,}164)(472{,}000)} \approx 88{,}000$$

The 90-percent confidence interval as shown by the data is from 331,000 to 613,000. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all samples.

Standard errors of estimated percentages. This section refers to percentages of a group of persons, families, or households possessing a particular attribute.

The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more, e.g., the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameters for the numerator. The approximate standard error, $s_{x,p}$, of the estimated percentage p can be obtained by the formula

$$s_{x,p} = \sqrt{\frac{b}{x}\left[p(100-p)\right]} \qquad (2)$$

Here x is the size of the subclass of households or persons in households which is the base of the percentage, p is the percentage (0<p<100), and b is the "b" parameter for the numerator.

1 Sampling Techniques, 3rd Ed. (New York: John Wiley and Sons, 1977), p. 321.

Illustration of the computation of the standard error of an estimated percentage. Suppose that, in July, of the 16,812,000 persons in nonfarm households with a mean monthly household cash income of $4,000 to $4,999, 6.7 percent were black. Using formula (2) and the "b" parameter of 9343 and a factor of 1.3333 for July from table 3, the approximate standard error is

$$\sqrt{\frac{(12,457)}{(16,812,000)}(6.7)(100-6.7)} \approx 0.7 \text{ percent}$$

Consequently, the 90 percent confidence interval as shown by these data is from 5.6 to 7.8 percent.

Standard error of a difference. The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2} \qquad (3)$$

where s_x and s_y are the standard errors of the estimates x and y. The estimates can be numbers, percents, ratios, etc. The above formula assumes that the sample correlation coefficient, r, between the two estimates is zero. If r is really positive (negative), then this assumption will lead to overestimates (underestimates) of the true standard error.

Illustration of the computation of the standard error of a difference. Suppose that SIPP estimates show the number of persons age 35-44 years in nonfarm households with mean monthly household cash income of $4,000 to $4,999 during the third quarter of 1985 was 3,186,000 and the number of persons age 25-34 years in nonfarm households with mean monthly household cash income of $4,000 to $4,999 in the same time period was 2,619,000. The standard errors of these numbers are 163,000 and 148,000, respectively. Assuming that these two estimates are not correlated, the standard error of the estimated difference of 567,000 is

$$\sqrt{(163,000)^2 + (148,000)^2} \approx 220,000$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons with mean monthly household cash income of $4,000 to $4,999 during the third quarter of 1985, (X), was different for persons age 35-44 years in nonfarm households than for persons age 25-34 years in nonfarm households. The difference, $X_{35-44} - X_{25-34}$ is 567,000. The difference divided by the standard error of the difference, $(X_{35-44} - X_{25-34})/s_{DIFF}$, is 2.58. Since the ratio is greater than 1.6, the data show that the difference between the two age groups is significant at the 10 percent level.

Standard error of a mean. A mean is defined here to be the average quantity of some item (other than persons, families, or households) per person, family, or household. For example, it could be the average monthly household income of females age 25 to 34. The standard error of a mean can be approximated by formula (4) below. Because of the approximations used in developing formula (4), an estimate of the standard error of the mean obtained from that formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean $\bar{x}$ is

$$s_{\bar{x}} = \sqrt{\frac{b}{y} s^2} \qquad (4)$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

The estimated population variance, s^2, is given by formula (5):

$$s^2 = \sum_{i=1}^{c} p_i \; x_i^2 - \overline{x}^2 \qquad (5)$$

where it is assumed that each person or other unit was placed in one of c groups; p_i is the estimated proportion of group i; $x_i = (Z_{i-1} + Z_i)/2$ where Z_{i-1} and Z_i are the lower and upper interval boundaries, respectively, for group i. The estimate x_i is assumed to be the most representative value for the characteristic of interest in group i. If group c is open-ended, i.e., no upper interval boundary exists, then an approximate value for x_c is

$$x_c = \frac{3}{2} Z_{c-1} \qquad (6)$$

Illustration of the Computation of the Standard Error of an Estimated Mean. Suppose that the average of monthly household incomes during the third quarter 1985 of persons age 25 to 34 are given in the following table.

Table 2. Distribution of Monthly Household Income Among Persons 25 To 34 Years Old.

	Total	under $300	$300 to $599	$600 to $899	$900 to $1,199	$1,200 to $1,499	$1,500 to $1,999	$2,000 to $2,499	$2,500 to $2,999	$3,000 to $3,499	$3,500 to $3,999	$4,000 to $4,999	$5,000 to $5,999	$6,000 and over
Thousands in interval	39,851	1371	1651	2259	2734	3452	6278	5799	4730	3723	2519	2619	1223	1493
Percent with at least as much as lower bound of interval	..	100.0	96.6	92.4	86.7	79.9	71.2	55.5	40.9	29.1	19.7	13.4	6.8	3.7

Using formula (5) and the mean monthly household cash income of $2,530 the approximate population variance, s^2, is

$$s^2 = \frac{1,371}{39,851}(150)^2 + \frac{1,651}{39,851}(450)^2 + \ldots + \frac{1,493}{39,851}(9,000)^2 - (2,530)^2$$

$$= 3,159,887.$$

Using formula (4), an appropriate "b" parameter of 6944 from table 3 and the factor 1.2222 for the third quarter of 1985, the estimated standard error of a mean $\overline{x}$ is

$$s_{\bar{x}} = \sqrt{\frac{8,487}{39,851,000}(3,159,887)^2} = \$26$$

Standard error of a median. The median quantity of some items such as income for a given group of persons, families, or households is that quantity such that at least half the group have as much or more and at least half the group have as much or less. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68 percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using formula (2), the standard error of an estimate of 50 percent of the group;

2. Add to and subtract from 50 percent the standard error determined in step (1);

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group owning more is equal to the smaller percentage found in step (2). This quantity will be the upper limit for the 68 percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group owning more is equal to the larger percentage found in step (2). This quantity will be the lower limit for the 68 percent confidence interval;

4. Divide the difference between the two quantities determined in step (3) by two to obtain the standard error of the median.

To perform step (3), it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. We recommend Pareto interpolation in most instances. Interpolation is used as follows. The quantity of the item such that "p" percent own more is

$$X_{pN} = \exp\left[\frac{\mathrm{Ln}\ (pN/N_1)}{\mathrm{Ln}\ (N_2/N_1)}\ \mathrm{Ln}\ (A_2/A_1)\right] A_1 \qquad (7)$$

if Pareto interpolation is indicated and

$$X_{pN} = \frac{pN-N_1}{N_2-N_1}\ (A_2-A_1) + A_1 \qquad (8)$$

if linear interpolation is indicated, where

N is size of the group,
A_1 and A_2 are the lower and upper bounds, respectively, of the interval in which X_{pN} falls,
N_1 and N_2 are the estimated number of group members owning more than A_1 and A_2, respectively,
exp refers to the exponential function, and
Ln refers to the natural logarithm function.

It should be noted that a mathematically equivalent result is obtained by using common logarithms (base 10) and antilogarithms.

Illustration of the Computation of a Confidence Interval and the Standard Error for a Median. To illustrate the calculations for the sampling error on a median, we return to the same example used to illustrate the standard error of a mean. The median monthly income for this group is $2,158. The size of the group is 39,851,000.

1. Using formula (2), the standard error of 50 percent on a base of 39,851,000 is about 0.7 percentage points.

2. Following step (2), the two percentages of interest are 49.3 and 50.7.

3. By examining table 2, we see that the percentage 49.3 falls in the income interval from \$2,000 to \$2,499. (Since 55.5 percent receive more than \$1,999 per month, but only 40.9 percent receive more than \$2,499 per month, the quantity that exactly 49.3 percent receive more than must be between \$2,000 and \$2,499.) Thus A_1 = \$2,000, A_2 = \$2,500, N_1 = 22,106,000, and N_2 = 16,307,000. Implementing Pareto interpolation, the upper bound of a 68 percent confidence interval for the median is

$$\exp\left[\frac{\operatorname{Ln}\dfrac{(.493)(39{,}851{,}000)}{22{,}106{,}000}}{\operatorname{Ln}\dfrac{16{,}307{,}000}{22{,}106{,}000}}\operatorname{Ln}\frac{2{,}500}{2{,}000}\right]\$2{,}000 = \$2{,}181$$

Also by examining table 2, we see that the percentage 50.7 falls in the same income interval. Thus, A_1, A_2, N_1, and N_2 are the same. So the lower bound of a 68 percent confidence interval for the median is

$$\exp\left[\frac{\operatorname{Ln}\dfrac{(.507)(39{,}851{,}000)}{22{,}106{,}000}}{\operatorname{Ln}\dfrac{16{,}307{,}000}{22{,}106{,}000}}\operatorname{Ln}\frac{2{,}500}{2{,}000}\right]\$2{,}000 = \$2{,}136$$

Thus, the 68 percent confidence interval on the estimated median is from \$2,136 to \$2,181. An approximate standard error is

$$\frac{\$2{,}181 - \$2{,}136}{2} = \$23.$$

Using linear interpolation, the 68 percent confidence interval of the estimated median is \$2,164 to \$2,212 and the approximate standard error is \$24.

Standard errors of ratios of means and medians. The standard error for a ratio of means or medians is approximated by formula (9):

$$s_{\left[\frac{x}{y}\right]} = \sqrt{\left[\frac{x}{y}\right]^2 \left[\left[\frac{s_y}{y}\right]^2 + \left[\frac{s_x}{x}\right]^2\right]} \qquad (9)$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula (9) assumes that the means or medians are not correlated. If the correlation between the two means or medians is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error for the ratio of means and medians.

TABLE 3. SIPP 1984 Generalized Variance Parameters for the Wave 7 Public Use File

PERSONS[1]	a	b
Total or White		
16+ Program Participation and Benefits, Poverty (3)		
Both Sexes	-0.0001144	20,370
Male	-0.0002404	20,370
Female	-0.0002182	20,370
16+ Income and Labor Force (4)		
Both Sexes	-0.0000390	6,944
Male	-0.0000819	6,944
Female	-0.0000744	6,944
All Others[2] (5)		
Both Sexes	-0.0001082	25,255
Male	-0.0002233	25,255
Female	-0.0002097	25,255
Black		
Poverty (1)		
Both Sexes	-0.0006186	17,372
Male	-0.0013259	17,372
Female	-0.0011595	17,372
All Others (2)		
Both Sexes	-0.0003327	9,343
Male	-0.0007131	9,343
Female	-0.0006236	9,343
HOUSEHOLDS/Families/Unrelated Individuals		
Total or White	-0.0000993	8,582
Black	-0.0006246	5,929

Factors to be Applied to Base Parameters to Obtain Parameters for Specific Reference Periods	
May 1985	4.0000
June	2.0000
July	1.3333
August	1.0000
September	1.3333
October	2.0000
November	4.0000
3rd Quarter 1985	1.2222

1 For cross-tabulations, use the parameters of the characteristic with the smaller number in parentheses.

2 For example, use these parameters for retirement and pension tabulations, 0+ program participation, 0+ benefits, 0+ income, and 0+ labor force tabulations, in addition to any other types of tabulations not specifically covered by another characteristic in this table.

Table 4. Metropolitan Subsample Factors (Multiply these factors by the weight for the person, family or household)

		Factors for use in State or MSA Tabulations	Factors for use in Regional or National Tabs
Northeast:	Connecticut	1.0390	1.0432
	Maine	--	--
	Massachusetts	1.0000	1.0040
	New Jersey	1.0000	1.0040
	New York	1.0110	1.0150
	Pennsylvania	1.0025	1.0065
	Rhode Island	1.2549	1.2599
Midwest:	Illinois	1.0232	1.0310
	Indiana	1.0000	1.0076
	Iowa	--	--
	Kansas	1.6024	1.6146
	Michigan	1.0000	1.0076
	Minnesota	1.0000	1.0076
	Missouri	1.0611	1.0692
	Nebraska	1.7454	1.7587
	Ohio	1.0134	1.0211
	Wisconsin	1.0700	1.0782
South:	Alabama	1.1441	1.1511
	Arkansas	1.0000	1.0061
	Delaware	1.0000	1.0061
	D.C.	1.0000	1.0061
	Florida	1.0333	1.0396
	Georgia	1.0000	1.0061
	Kentucky	1.1124	1.1192
	Louisiana	1.1470	1.1540
	Maryland	1.0000	1.0061
	North Carolina	1.0000	1.0061
	Oklahoma	1.1146	1.1214
	South Carolina	1.1270	1.1339
	Tennessee	1.0000	1.0061
	Texas	1.0192	1.0254
	Virginia	1.0778	1.0844
	West Va.-Miss.	--	--
West:	Arizona	1.0870	1.0870
	California	1.0000	1.0000
	Colorado	1.0000	1.0000
	Hawaii	1.0000	1.0000
	Oregon	1.0879	1.0879
	Washington	1.0868	1.0868

-- indicates no metropolitan subsample is shown for the State.